531,195

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
29 April 2004 (29.04.2004)

PCT

(10) International Publication Number
WO 2004/036919 A1

(51) International Patent Classification[7]: H04N 7/26

(21) International Application Number:
PCT/IB2003/004452

(22) International Filing Date: 8 October 2003 (08.10.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/418,961 16 October 2002 (16.10.2002) US
60/483,796 30 June 2003 (30.06.2003) US

(71) Applicant *(for all designated States except US)*: **KONINKLIJKE PHILIPS ELECTRONICS N.V.** [NL/NL]; Groenewoudseweg 1, NL-5621 BA Eindhoven (NL).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **YE, Jong, Chul** [US/US]; P.O. Box 3001, Briarcliff Manor, NY 10510-8001 (US). **VAN DER SCHAAR, Mihaela** [US/US]; P.O. Box 3001, Briarcliff Manor, NY 10510-8001 (US).

(74) Common Representative: **KONINKLIJKE PHILIPS ELECTRONICS N.V.**; c/o Russell Gross, P.O. Box 3001, Briarcliff Manor, NY 10510-8001 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

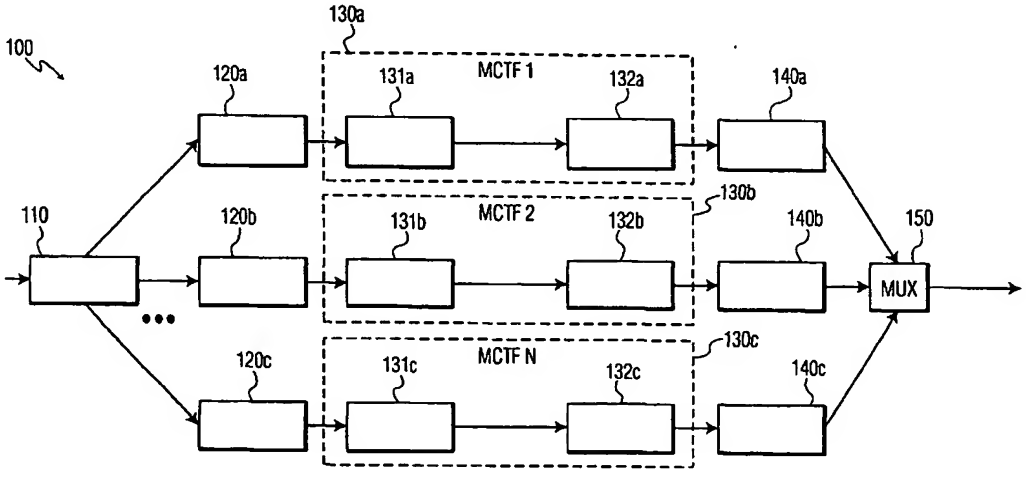
(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declaration under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE,*

[Continued on next page]

(54) Title: FULLY SCALABLE 3-D OVERCOMPLETE WAVELET VIDEO CODING USING ADAPTIVE MOTION COMPENSATED TEMPORAL FILTERING



(57) Abstract: A method and device for coding video where a video signal is spatially decomposed into at least two signals of different frequency sub-bands, an individualized motion compensated temporal filtering scheme is applied to each sub-band signal adaptively according to signal contents, and texture coding is applied to each of the motion compensated temporally filtered subband signals adaptively according to the signal content.

WO 2004/036919 A1

DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

Published:

— *with international search report*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

FULLY SCALABLE 3-D OVERCOMPLETE WAVELET VIDEO CODING USING ADAPTIVE MOTION COMPENSATED TEMPORAL FILTERING

This application claims the benefit under 35 USC 119(e) of U.S. provisional application serial no. 60/418,961, filed on October 16, 2002, which is incorporated herein by reference.

The present invention relates to video compression, and more particularly to overcomplete wavelet video coding using adaptive motion compensated temporal filtering.

Current video coding algorithms are mainly based on hybrid-coding schemes with motion compensated predictive coding. In such hybrid schemes, temporal redundancy is reduced using motional compensation and spatial resolution is reduced by transform coding the residue of motion compensation. These hybrid-coding schemes, however, are prone to error propagation and lack flexibility in terms of providing true scalable bitstream, i.e., the ability to decompress to different quality, resolution, and frame-rate layers from the same compressed bitstream.

In contrast, 3D sub-band/wavelet coding can provide very flexible scalable bitstream and higher error resilience. Wavelet-based scalable video coding schemes permit great flexibility in terms of the different scalability types allowed. Hence, they are especially useful for video transmission over heterogeneous wireless and wired networks, to various devices with different capabilities.

Currently, there are two wavelet-based video coding schemes: overcomplete wavelet and interframe wavelet. In overcomplete (OW) wavelet video coding, the spatial wavelet transform for each frame is performed first, followed by exploitation of interframe redundancy by predicting the wavelet coefficient values, or by defining temporal contexts in entropy coding. In interframe wavelet video coding, wavelet filtering is performed along the temporal axis followed by a 2D spatial wavelet transform.

Present interframe wavelet video coding schemes use motion compensated temporal filtering (MCTF), to reduced the temporal redundancy. MCTF is performed in the temporal direction of motion before spatial decomposition is performed. Such video coding schemes are referred to herein as spatial domain MCTF (SDMCTF). However, the quality of the matches provided by the motion estimation algorithm inherently limit SDMCTF video coding schemes. For example, some of the interframe wavelet-coded sequence appears to be slightly blurred, because imperfect motion estimation causes movement of frame details

into the temporal high frequency sub-bands, and from there, to spatial high frequency sub-bands. These artifacts lead to degraded visual performance for unquantized and spatially scaled sequences. Further tests have indicated that decreasing the number of temporal decomposition levels can reduce the artifacts.

In present OW video coding schemes, wavelet filtering is used to spatially decompose each of the video frames into multiple sub-bands, and temporal correlation for each sub-band is removed using motion estimation.

There have been many attempts to predict the wavelet coefficients by motion compensation in the wavelet domain. However, motion compensation in the wavelet domain is highly dependent on the alignment of the signal and the discrete grid chosen for the analysis. There exist very large differences between the wavelet coefficients of the original image and the one-pixel-shifted image. This shift-variant property happens frequently around the image edges, so motion compensation of the wavelet coefficients can be difficult.

Existing OW video coding schemes overcome the inefficiency of motion estimation in wavelet domain by utilizing the odd-phase wavelet coefficients in the prediction as well. A convenient method of obtaining the odd phase coefficients is to perform band shifting. Since the decoded previous frame is also available at the decoder, prediction from overcomplete expansion does not require any additional overhead. Moreover, the computational complexity of searching both optimal phase and motion vectors in wavelet domain is comparable to that of conventional motion estimation in spatial domain with fractional pel accuracy.

However, due to the motion estimation/compensation, the conventional OW framework suffers from drift, which results in performance loss in SNR scalability. Furthermore, only limited range of temporal scalability can be achieved using B frames.

Accordingly, a wavelet-based video-coding scheme with improved SNR and temporal scalability is needed.

The present invention is directed to a method and device for coding video. According to a first aspect of present invention, a video signal is spatially decomposed into at least two signals of different frequency sub-bands. An individualized motion compensated temporal filtering scheme is applied to each sub-band signal. Texture coding is then applied to each of the motion compensated temporally filtered subband signals. According to a second aspect of the invention, a signal including at least two encoded

motion compensated temporally filtered, different frequency sub-band signals of a video signal, is decoded. Inverse motion compensated temporal filtering is independently applied to each of the decoded at least two sub-band signals. The at least two sub-band signals are spatially recomposed and the video signal is reconstructed from at least one of the at least two spatially recomposed sub-band signals.

FIG. 1 is a block diagram of a 3-D overcomplete wavelet video encoder according to an exemplary embodiment of the present invention, which may be used for performing the IBMCTF method of present invention.

FIG. 2 is a block diagram of an adaptive higher order interpolation filter used in the present invention.

FIG. 3 illustrates the generation of an extended reference frame for motion estimation from overcomplete expansion of wavelet coefficients according to the present invention.

FIG. 4A illustrates a decomposition scheme for conventional MCTF that generates blurred images.

FIG. 4B illustrates a decomposition scheme used in the present invention.

FIG. 5 is a block diagram of a 3-D overcomplete wavelet video decoder according to an exemplary embodiment of the present invention.

FIG. 6 shows an over-complete wavelet expansion using a LBS algorithm for two level decompositions.

FIG. 7 is a video of a 2-level overcomplete wavelet transform obtained using the LBS method.

FIG. 8 illustrates the interleaving scheme of the present invention for a 1-D case of a one level decomposition.

FIG. 9 shows the overcomplete wavelet coefficients of the first frame of the video of FIG. 7 after performing the interleaving process of the present invention.

FIG. 10 is a wavelet block form by the LBS algorithm.

FIG. 11 shows a Table that illustrates the MAD in wavelet domain for temporal high sub-band frames.

FIGS. 12-17 plot the rate distortion performance of the IBMCTF video coding scheme of the present invention and SDMCTF for several test sequence for integer and 1/8-pel accurate motion estimation.

FIG. 18 is an exemplary embodiment of a system which may be used for implementing the principles of the present invention.

The present invention is a fully scalable three-dimensional (3-D) overcomplete wavelet video coding scheme that utilizes a novel inband motion compensated temporal filtering (IBMCTF) method. The IBMCTF method of the present invention overcomes the drawbacks of previous IBMCTF coding methods, and demonstrates coding efficiency comparable or better than conventional interframe wavelet coding methods that utilize spatial domain motion compensated temporal filtering.

FIG. 1 is a block diagram of a 3-D overcomplete wavelet video encoder according to an exemplary embodiment of the present invention, which may be used for performing the IBMCTF method of present invention. The video encoder 100 includes a 3-D wavelet transform unit 110 that spatially decomposes each video frame of an input video into any desired number of multiple sub-bands 1, 2,... and N using a conventional 3-D overcomplete wavelet filtering process.

The video encoder 100 further includes a partitioning unit 120a, 120b, 120c for each sub-band generated by the wavelet transform unit 110. Each partitioning unit 120a, 120b, 120c divides the wavelet coefficients of its associated sub-band into groups of frames (GOFs) for encoding as a group.

The video encoder 100 also includes a motion compensated temporal filtering (MCTF) unit 130a, 130b, 130c for each sub-band, that contains a motion estimator 131a, 131b, 131c and a temporal filter 132a, 132b, 132c. Each MCTF 130a, 130b, 130c separately removes temporal correlation or redundancy from the GOFs of each sub-band using a motion compensated temporal filtering (MCTF) process. In accordance with the present invention, the use of a discrete MCTF unit for each sub-band allows the motion compensated temporal filtering process to be tailored for each sub-band independently of the other sub-bands. In addition, the temporal filtering process selected for a particular sub-band may be based on different criteria.

The encoder additionally includes a texture encoder 140a, 140b, 140c for each sub-band that allows the residual signal and motion information (motion vectors) generated by the MCTF units 130a, 130b, 130c for each sub-band to be independently texture coded using any optimized texture coding process. The texture coded residual signals and motion information are then combined into a single bitstream by a multiplexer 150. Another embodiment of texture coding is a gobal transform of a full size residual frame, which is

applied after the all residual signals and motion information generated by the MCTF units 130a, 130b, 130c for each sub-band are combined to generate the full size residual frame.

As one of ordinary skill in the art will appreciate, the critical-sampled wavelet decomposition in known IBMCTF methods is only periodically shift-invariant. Therefore, performing motion estimation and compensation in the wavelet domain is inefficient and may incur a coding penalty. To address this problem, each motion compensated filtering unit 130a, 130b, 130c utilizes an adaptive higher order interpolation filter 200, as shown in FIG. 2, to maximize the performance of the motion estimator 131a, 131b, 131c. The interpolation filter 200 of the invention includes a low band shifting (LBS) unit 210 that performs low band shifting, an interleaving unit 220 that performs overcomplete wavelet coefficient interleaving, and an interpolation unit 230. The LBS process is implemented in the LBS unit 210 with one or more known LBS algorithms that efficiently generate an overcomplete representation of the original wavelet coefficients, which is now shift invariant. LBS advantageously generates the overcomplete expansion of the original wavelet coefficients at the encoder and decoder using one or more similar LBS algorithms, therefore, no additional information needs to be encoded and transmitted as compared to conventional interframe wavelet coding schemes.

The interleaving process, performed by the interleaving unit 220, combines the different phase information provided by the overcomplete wavelet coefficients to generate an extended reference frame. Accordingly, there is no need to encode the phase information separately as in previous IBMCTF based video coding methods. Due to the interleaving process of the present invention, the phase information is coded inherently as part of the higher accuracy motion vectors.

From the extended reference frame, the interpolation unit 230 generates a fractional pel, such as 1/2, 1/4, 1/8, 1/16 pels, which is used by the motion estimator 131a, 131b, 131c for motion estimation. Interpolation may be implemented with a conventional one-dimensional interpolation filter. In order to maximize the performance of the motion estimation and MCTF, independently optimized interpolation filters with a different tap can be used for each subband. FIG. 3 illustrates the generation of an extended reference frame for motion estimation from overcomplete expansion of wavelet coefficients according to the present invention. In order to achieve a higher order interpolation for motion estimation in the HH sub-band overcomplete expansion 300, for example, three other phases of wavelet coefficients are generated from original wavelet coefficients 310 by shifting the lower sub-

band with the amount of (1,0), (0,1) and (1,1). Then, four phases of wavelet coefficients 310, 320, 330, 340 are interleaved to generate an extended reference frame 350.

The IBMCTF based 3-D overcomplete wavelet video coding method of the present invention provides improved spatial scalability performance as compared with known spatial domain motion compensated temporal filtering (SDMCTF) based video coding methods. This is because the temporal filtering is performed per sub-band (resolution) and hence, loss of information from the finer resolution sub-bands does not incur any drift in the temporal direction.

As mentioned earlier, the use of a discrete MCTF unit 130a, 130b, 130c for each sub-band allows different temporal filtering techiques to be used at the various resolutions. For example, in one embodiment, a bi-directional temporal filtering technique can be used for low resolution sub-bands, while a forward temporal filtering technique can be used for higher resolution sub-bands. The temporal filtering technique can be selected based on minimizing a distortion or a complexity measure (e.g. the low resolution sub-bands have less pixels and hence bi-directional and multiple reference temporal filtering can be employed, while for the high resolution sub-bands that have a larger number of pixels, only forward estimation is performed). Such a flexible choice of temporal filtering options makes moves the invention away from the strict 1D+2D decomposition schemes as performed by MCTF, to a more general 3-D decomposition scheme with spatial size reduction throughout the temporal levels, where the higher spatial frequency sub-bands are omitted from longer-term temporal filtering.

The use of a discrete partitioning unit 120a, 120b, 120c for each sub-band allows the GOFs to be adaptively determined per sub-band. For instance, the LL-sub-bands might have a very large GOF, while the H-sub-bands can use limited GOFs. The GOF sizes can be varied based on the sequence characteristics, complexity or resiliency requirements. As mentioned earlier, the decomposition scheme for conventional MCTF, as shown in FIG. 4A, generates blurred images. However, the use of different temporal decomposition levels and GOF sizes allows the 3-D wavelet scalable video coding scheme of the present invention to overcomes such drawbacks. As shown in FIG. 4B, GOF sizes for LL LH (HL), and HH may be 8, 4, and 2 frames, respectively, which allow maximum decomposition levels of 3, 2, and 1 respectively. This way the higher spatial frequency sub-bands are omitted from longer-term temporal filtering.

The number of temporal decomposition levels for the various sub-bands can be determined either based on content, or to reduce a specific distortion metric or simply based on the desired temporal scalability in each resolution. For instance, if 30, 15 and 7.5 Hz frame-rates are desired at CIF (352x288) size resolution, and only 30 and 15 at SD (704x576) size resolution, then for the LL spatial-sub-band, three levels of temporal decomposition are used, while only two levels of temporal decomposition can be applied for LH, HL, and the HH sub-bands.

As also mentioned earlier, the use of discrete texture coding unit 140a, 140b, 140c for each sub-band allows adaptive texture coding of the various spatial sub-bands. For example, wavelet or DCT-based texture coding schemes may be used. If DCT-based texture coding is used, intra-coded blocks can be advantageously inserted anywhere within the GOF to deal efficiently with covering and uncovering situations. Also, "adaptive intra-refresh" concepts from MPEG-4/H.26L can be easily employed to provide improved resiliency, and different refresh rates can be used for the various sub-bands to obtain different resiliencies. This is especially beneficial since the lower resolution sub-bands can be used for concealing the higher resolution sub-bands and hence, their resiliency is more important.

Another advantage of the present invention relates to the complexity scalability of the decoder. If there are many decoders with different computation power and displays, the same scalable bitstream can be used to support all those decoders through SNR/spatial/temporal scalability. For example, the scalable bitstream generated by the encoder of the invention can be decoded with a decoder with low complexity that can decode only low resolution spatial and temporal decomposition level, which incurs only small computational burden. Similarly, the scalable bitstream generated by the encoder of the invention can also be decoded with a decoder having sophisticated decoding power that can decode the whole bit stream to achieve the full spatial and temporal resolution.

FIG. 5 is a block diagram of a 3-D overcomplete wavelet video decoder according to an exemplary embodiment of the present invention. The decoder may be used for decoding the bitstream produced by the encoder of the present invention. The video decoder 400 may include a demultiplexer 410 that processes the bitstream to separate the encoded wavelet coefficients from the motion information.

A first texture decoder 420 texture decodes the wavelet coefficients, according to the inverse of the texture coding technique performed on the encoding side, into their separate sub-bands 1, 2,...and N. The wavelet coefficients of a sub-band produced by the first

texture decoder 420 correspond to each GOF of that sub-band. A motion vector decoder 430 decodes the motion information for each sub-band according to the inverse of the texture coding technique performed on the encoding side. Using the decoded motion vectors and residual texture information, inverse MCTF is applied by MCTF units 440a, 440b, 440c for each sub-band independently and an inverse wavelet transform unit 450 spatially recomposes each sub-band to reconstruct the low, medium, and high level images. The low-band-shifting block reads the recomposed sub-band images to assemble a full size image and then the low band shifted wavelet decomposition is applied to provide the extended reference frames for the inverse MCTF units 440a, 440b, 440c. Depending on the display resolution, a video reconstruction unit (not shown) may use one of the sub-bands to generate the low resolution video, or use two sub-bands to generate a medium resolution video, or use all of the sub-bands to generate a high resolution, full quality video.

The various processes utilized in the video scheme of the present invention will now be described in greater detail below.

MOTION ESTIMATION AND COMPENSATION IN THE OVERCOMPLETE WAVELET DOMAIN

1. Low Band Shifting Method (LBS)

The decimation process performed in a wavelet transform generates wavelet coefficients that are no longer shift-invariant. Hence, translation motion in the spatial domain cannot be accurately estimated from the wavelet coefficients, which in turn produces a significant loss in coding efficiency. The LBS algorithms utilized in the present invention provide a method for overcoming the shift-variant property of the wavelet transform. At a first level, the original and shifted signals are decomposed into low-sub-band and high-sub-band signals. Subsequently, the low-sub-band signal is further decomposed in the same way as for the first level.

FIG. 6 shows an over-complete wavelet expansion using a LBS algorithm for two level decompositions. The one dimensional (1-D) formulation can be easily expanded to wavelet decompositions having multiple levels and also to two-dimensional (2-D) image signals. The pair (m,n) indicates that the wavelet coefficients within that sub-band were generated by shifts of m-pixels in the x-direction and n-pixels in the y-direction, respectively. The LBS algorithm generates a full-set of wavelet coefficients for all the possible shifts of the input sub-band. Hence, the representation accurately conveys any shift

in spatial domain. As will be discussed further on, the different shifted wavelet coefficients corresponding to the same decomposition level at a specific spatial location are referred to as "cross-phase" wavelet coefficients.

FIG. 7 is a video of a 2-level overcomplete wavelet transform obtained using the LBS method. Note that for an n-level decomposition, the overcomplete wavelet representation requires a storage space that is 3n+1 larger than that of the original image.

2. Interleaving of Wavelet Coefficients

The novel interleaving scheme of the present invention stores the overcomplete wavelet coefficients differently from that depicted in FIGS. 6 and 7. As shown in FIG. 8, which illustrates the interleaving scheme of the present invention for a 1-D case of a one level decomposition, the coefficients for shift-interleaving is performed such that the new coordinates in the overcomplete domain correspond to the associated shift in the original spatial domain.

The interleaving scheme can be used recursively at each decomposition level and can be directly extended for 2-D signals. FIG. 9 shows the overcomplete wavelet coefficients of the first frame of the video of FIG. 7 after performing the interleaving process of the present invention. As can be seen from FIG. 9, the interleaved low sub-band signal is a low-pass filtered version of the original frame using the overcomplete wavelet low-pass filter. The interleaving process of the present invention enables the IBMCTF method of the present invention to provide sub-pixel accuracy motion estimation and compensation. Previously proposed IBMCTF schemes cannot provide optimal sub-pixel accuracy motion estimation and compensation, because they do not take into consideration cross-phase dependencies between neighbouring wavelet coefficients. Furthermore, the interleaving process allows the IBMCTF method of the invention to use hierarchical variable size block matching, backward motion compensation, and adaptive insertion of intra blocks.

Generation of Wavelet Block

As is well known in the art, in a wavelet decomposition, every coefficient at a given scale, with the exception of those in the highest frequency sub-bands, can be related to a set of coefficients of the same orientation at finer scales. In many wavelet coders, this relationship is exploited by representing the coefficients as a data structure called a wavelet

tree. In the LBS algorithm, the coefficients of each wavelet tree rooted in the lowest sub-band are rearranged to form a wavelet block, as shown in FIG. 10. The purpose of the wavelet block is to provide a direct association between the wavelet coefficients and what they represent spatially in the image. Related coefficients at all scales and orientations are included in each block.

Structure of Motion Estimation

In the spatial domain, the block-based motion estimation usually divides an image into small blocks and then finds the block of the reference frame that minimizes the mean absolute different (MAD) to each block of the current frame. The motion estimation of the LBS algorithm finds the motion vector (dx, dy) that generates the minimum MAD between the current wavelet block and the reference wavelet block. As an example, if an input image is decomposed up to the third level (i.e. the input image can be decomposed to a total of ten sub-bands), and the displacement vector is (dx,dy), then the MAD of the k-th wavelet block in FIG. 10 is computed as follows:

$$
\begin{aligned}
MAD_k(dx,dy) = \sum_{l=1}^{3} \sum_{x_l=x_{l,k}}^{x_{l,k}+M/2^l} \sum_{y_l=y_{l,k}}^{y_{l,k}+N/2^l} \{ \\
\left| HL_{cur}^{(l)}(x_l,y_l) - HL_{ref}^{(l)}(dx\%2^l, dy\%2^l; x_l + \left\lfloor \frac{dx}{2^l} \right\rfloor, y_l + \left\lfloor \frac{dy}{2^l} \right\rfloor) \right| \\
+ \left| LH_{cur}^{(l)}(x_l,y_l) - LH_{ref}^{(l)}(dx\%2^l, dy\%2^l; x_l + \left\lfloor \frac{dx}{2^l} \right\rfloor, y_l + \left\lfloor \frac{dy}{2^l} \right\rfloor) \right| \\
+ \left| HH_{cur}^{(l)}(x_l,y_l) - HH_{ref}^{(l)}(dx\%2^l, dy\%2^l; x_l + \left\lfloor \frac{dx}{2^l} \right\rfloor, y_l + \left\lfloor \frac{dy}{2^l} \right\rfloor) \right| \} \\
+ \sum_{x_l=x_{3,k}}^{x_{3,k}+M/2^3} \sum_{y_l=y_{3,k}}^{y_{3,k}+N/2^3} \left| LL_{cur}^{(3)}(x_l,y_l) - LL_{ref}^{(3)}(dx\%2^3, dy\%2^3; x_l + \left\lfloor \frac{dx}{2^3} \right\rfloor, y_l + \left\lfloor \frac{dy}{2^3} \right\rfloor) \right|
\end{aligned}
$$

where $x_{l,k} = x_{0,k}/2^l$ and $y_{l,k} = y_{0,k}/2^l$; and $(x_{0,k}, y_{0,k})$ denotes the initial position of the k-th wavelet block in the spatial domain, as shown in FIG. 10 and $\lfloor x \rfloor$ denotes largest integer not bigger than x. Here, for example, the i-th level HL sub-band of the reference frame is represented by $HL_{ref}^{(i)}(m,n;x,y)$, where (m,n) denotes the number of shift in x- and y-direction in the spatial domain and (x,y) is the location of the sub-band signal. The optimization criterion for the motion estimation is now finding the optimal (dx,dy) which minimizes this MAD. Note that in the original LBS algorithm, for the non-integer value of (dx,dy), it is not possible to compute the MAD using the above formula. More specifically, the MAD in conventional IBMCTF video coding schemes is based solely on the same-

phase wavelet coefficients and the resulting sub-pixel accuracy motion estimation and compensation is not optimal.

However, in the IBMCTF method of the present invention, the interleaving process enables the MAD calculation to be performed similarly as in SDMCTF video coding schemes, even for the sub-pixel accuracy. More specifically, the MAD for the displacement vector (dx,dy) for the IBMCTF method of the present invention is computed as follows:

$$MAD_k(dx,dy) = \sum_{l=1}^{3} \sum_{x_l=x_{l,k}}^{x_{l,k}+M/2^l} \sum_{y_l=y_{l,k}}^{y_{l,k}+N/2^l} \{$$
$$\left|HL_{cur}^{(l)}(x_l,y_l) - LBS\_HL_{ref}^{(l)}(2^l x_l + dx, 2^l y_l + dy)\right| + \left|LH_{cur}^{(l)}(x_l,y_l) - LBS\_LH_{ref}^{(l)}(2^l x_l + dx, 2^l y_l + dy)\right|$$
$$+ \left|HH_{cur}^{(l)}(x_l,y_l) - LBS\_HH_{ref}^{(l)}(2^l x_l + dx, 2^l y_l + dy)\right| \}$$
$$+ \sum_{x_l=x_{3,k}}^{x_{3,k}+M/2^l} \sum_{y_l=y_{3,k}}^{y_{3,k}+N/2^l} \left|LL_{cur}^{(l)}(x_l,y_l) - LBS\_LL_{ref}^{(l)}(2^l x_l + dx, 2^l y_l + dy)\right|$$

where, for example, $LBS\_HL_{ref}^{(l)}(x,y)$ denotes the extended HL sub-band of reference frame using interleaving process of the present invention. Note that even if (dx,dy) are non-integer values, the same interpolation technique used for SDMCTF can be easily used for each extended sub-band to generate the MAD for the non-integer displacement. Therefore, the IBMCTF video coding scheme of the present invention provides more efficient and indeed *optimal* sub-pixel motion estimation compared to the existing IBMCTF coding schemes. Also, in the IBMCTF video coding scheme of the present invention with the wavelet block structure does not incur any motion vector overhead because the number of the motion vector to be coded is the same as that of SDMCTF. Since the motion estimation is closely aligned with the residual coding, a more sophisticated motion estimation criterion (such as the entropy of the residual signal) may be used to improve the coding performance.

SIMULATION RESULTS

In order to verify that motion estimation and motion compensation in accordance with the present invention in the overcomplete wavelet domain yields lower residual energy in the wavelet domain, we use a one level temporal decomposition and compute the MAD for both IBMCTF and SDMCTF. Note that in interframe wavelet coding, the MAD is computed in the spatial-domain, but actually what needs to be minimized is the residual energy in the wavelet domain. FIG. 11 shows a Table that illustrates the MAD in wavelet domain for temporal high sub-band frames. The MAD values are averaged over the first 50 frames of temporal high sub-bands. For the SDMCTF cases, the corresponding MAD values

in wavelet domains are computed after the wavelet transform of the residual signal . Note that the MAD for the IBMCTF is always smaller than for SDMCTF, which indicates the possible coding gain of the IBMCTF video coding scheme of the present invention over SDMCTF.

FIGS. 12-17 plot the rate distortion performance of the IBMCTF video coding scheme of the present invention and SDMCTF for several test sequence for integer and 1/8-pel accurate motion estimation. The inband structure for MCTF was computed with a two level spatial decomposition performed by a Daubechies 9/7 filter, and four levels of decomposition were used for the temporal direction. The texture coding was performed with an EZBC algorithm described in the article entitled, Invertible Three-Dimensional Analysis/Synthesis System For Video Coding With Half-Pixel Accurate Motion Compensation, by S.T. Hsiang et al., VCIP 1999, SPIE Vol. 3653, pp. 537-546. Similar to SDMCTF, the sub-pixel motion estimation using 1/8 pel greatly improves the coding performance of the IBMCTF. The overall coding performance of the IBMCTF and SDMCTF is comparable. However, some sequences such as "Coastguard", "Silent" and "Stefan" exhibit a performance gain of up to 0.5dB, while for the "Mobile" sequence a 0.3dB performance degradation can be observed. Visually, the IBMCTF algorithm of the present invention is free of blocking artefacts of the motion estimation since the motion estimation and filtering is done in each sub-band and the boundary of the motion is filtered out using wavelet recomposition filter.

FIG. 18 is an exemplary embodiment of a system 500 which may be used for implementing the principles of the present invention. The system 500 may represent a television, a set-top box, a desktop, laptop or palmtop computer, a personal digital assistant (PDA), a video/image storage device such as a video cassette recorder (VCR), a digital video recorder (DVR), a TiVO device, etc., as well as portions or combinations of these and other devices. The system 500 includes one or more video/image sources 501, one or more input/output devices 502, a processor 503 and a memory 504. The video/image source(s) 501 may represent, e.g., a television receiver, a VCR or other video/image storage device. The source(s) 501 may alternatively represent one or more network connections for receiving video from a server or servers over, e.g., a global computer communications network such as the Internet, a wide area network, a metropolitan area network, a local area network, a terrestrial broadcast system, a cable network, a satellite network, a wireless

network, or a telephone network, as well as portions or combinations of these and other types of networks.

The input/output devices 502, processor 503 and memory 504 may communicate over a communication medium 505. The communication medium 505 may represent, e.g., a bus, a communication network, one or more internal connections of a circuit, circuit card or other device, as well as portions and combinations of these and other communication media. Input video data from the source(s) 501 is processed in accordance with one or more software programs stored in memory 504 and executed by processor 503 in order to generate output video/images supplied to a display device 506.

In a preferred embodiment, the coding and decoding principles of the present invention may be implemented by computer readable code executed by the system. The code may be stored in the memory 504 or read/downloaded from a memory medium such as a CD-ROM or floppy disk. In other embodiments, hardware circuitry may be used in place of, or in combination with, software instructions to implement the invention. For example, the functional elements shown in FIGS. 1, 2, and 5 may also be implemented as discrete hardware elements.

While the present invention has been described above in terms of specific embodiments, it is to be understood that the invention is not intended to be confined or limited to the embodiments disclosed herein. For example, other transforms besides DCT can be employed, including but not limited to wavelets or matching-pursuits. These and all other such modifications and changes are considered to be within the scope of the appended claims.

CLAIMS

1. A method of encoding video, the method comprising the steps of:

providing a video signal;

spatially decomposing (110) the video signal into at least two signals of different frequency sub-bands;

applying an individualized motion compensated temporal filtering scheme (130a, 130b, 130c) to each sub-band signal; and

texture coding (140a, 140b, 140c) each of the motion compensated temporally filtered subband signals.

2. The method according to claim 1, wherein the spatially decomposing step (110) is performed by wavelet filtering.

3. The method according to claim 1, wherein the video signal defines a plurality of frames, the spatially decomposing step (110) including spatially decomposing each of the frames of the video signal into the at least two signals of different frequency sub-bands.

4. The method according to claim 1, wherein prior to the step (130a, 130b, 130c) of applying a motion compensated temporal filtering scheme (130a, 130b, 130c), further comprising the step of breaking each of the sub-band signals into a signal representing a group of temporal frames having a certain content.

5. The method according to claim 4, wherein the individualized motion compensated temporal filtering scheme (130a, 130b, 130c) applied to each sub-band signal is individualized according to the content of the group of frames.

6. The method according to claim 1, wherein prior to the step of applying a motion compensated temporal filtering scheme, further comprising the step of breaking each of the sub-band signals into a signal representing a group of frames (120a, 120b, 120c), the number of the frames in at least one of the group of frames signals being adaptively determined.

7. The method according to claim 1, wherein the individualized motion compensated temporal filtering scheme (130a, 130b, 130c) applied to each sub-band signal is individualized according to a spatial resolution of the sub-band signal.

8. The method according to claim 1, wherein the step of applying an individualized motion compensated temporal filtering scheme (130a, 130b, 130c) to each sub-band signal is performed by using variable accuracy motion estimation, which is dependent of signal contents.

9. The method according to claim 1, wherein the individualized motion compensated temporal filtering scheme (130a, 130b, 130c) applied to each sub-band signal is individualized according to a temporal correlation of the sub-band signal.

10. The method according to claim 1, wherein the step of applying an individualized motion compensated temporal filtering scheme (130a, 130b, 130c) to each sub-band signal is performed by using an individualized interpolation filter (200) for maximizing motion estimation performance.

11. The method according to claim 1, wherein the individualized motion compensated temporal filtering scheme (130a, 130b, 130c) applied to each sub-band signal is individualized according to a characteristic of the sub-band signal.

12. The method according to claim 1, wherein the step of applying an individualized motion compensated temporal filtering scheme(130a, 130b, 130c) to each bandwidth signal is performed by using a temporal filter selected from the group consisting of multi-directional temporal filters and unidirectional temporal filters.

13. The method according to claim 1, wherein the step of applying an individualized motion compensated temporal filtering scheme (130a, 130b, 130c) to each sub-band signal includes the steps of:

shifting (210) the sub-band signal, which is from a phase of wavelet coefficients generated in the spatially decomposing step, at least three times to generate three additional phases of wavelet coefficients;

interleaving (220) the four phases of wavelet coefficients to produce an extended reference frame; and

estimating motion (131a, 131b, 131c) using the extended reference frame.

14. The method according to claim 13, wherein the spatial decomposing step (110) is performed to provide a plurality decomposition levels, each decomposition level comprising a different frequency sub-band and wherein the step of applying the individualized motion compensated temporal filtering scheme (130a, 130b, 130c), by performing the shifting (210), interleaving (220) and estimating steps 131a, 131b, 131c), is recursively applied for each decomposition level.

15. The method according to claim 1, wherein the step of applying an individualized motion compensated temporal filtering scheme (130a, 130b, 130c) to each sub-band signal includes the steps of:

shifting (210) the sub-band signal, which are from a phase of wavelet coefficients generated in the spatially decomposing step, at least three times to generate three additional phases of wavelet coefficients;

combining (220) the four phases of wavelet coefficients to produce an extended reference frame;

generating a fractional pel (230) from the extended frame; and

estimating motion (131a, 131b, 131c) according to the fractional pel.

16. The method according to claim 14, wherein the spatial decomposing step (110) is performed to provide a plurality decomposition levels, each decomposition level comprising a different frequency sub-band and wherein the step of applying the individualized motion compensated temporal filtering scheme(130a, 130b, 130c), by performing the shifting (210), combining (220), generating (230) and estimating steps (131a, 131b, 131c), is recursively applied for each decomposition level.

17. A memory medium for encoding video, the memory medium comprising:

code for spatially decomposing (110) a video signal into at least two signals of different frequency sub-bands;

code for applying an individualized motion compensated temporal filtering scheme (130a, 130b, 130c) to each sub-band signal; and

code for texture coding (140a, 140b, 140c) each of the motion compensated temporally filtered subband signals.

18. A device for encoding video, the device comprising:

a wavelet transform unit (110) for spatially decomposing a video signal into at least two signals of different frequency sub-bands;

a motion compensated temporal filtering unit (130a, 130b, 130c) for each of the at least two sub-band signals, each motion compensated temporal filtering unit applying an individualized motion compensated temporal filtering scheme to its associated sub-band signal; and

a texture coding unit (140a, 140b, 140c) for each of the at least two sub-band signals, each texture coding unit texture coding its associated motion compensated temporally filtered subband signal.

19. The device according to claim 18, further comprising a partitioning unit (120a, 120b, 120c) for each of the sub-band signals, each partitioning unit breaking its associated sub-band signal into a signal representing a group of temporal frames having a certain content.

20. The device according to claim 18, wherein each motion compensated temporal filtering unit (130a, 130b, 130c) includes:

a low band shifting unit (210) for shifting its associated sub-band signal, which is from a phase of wavelet coefficients, at least three times to generate three additional phases of wavelet coefficients; and

an interleaving unit (220) for interleaving the four phases of wavelet coefficients to produce an extended reference frame.

21. The device according to claim 20, wherein each motion compensated temporal filtering unit (130a, 130b,130c) further includes an interpolating unit (230) for generating a fractional pel from the extended frame.

22. The device according to claim 21, wherein each motion compensated temporal filtering unit (130a, 130b,130c) further includes a motion estimation unit (131a, 131b, 131c) for estimating motion according to the fractional pel.

23. A method of decoding video, the method comprising the steps of:

decoding (420) a signal including at least two encoded motion compensated temporally filtered, different frequency sub-band signals of a video signal;

independently applying inverse motion compensated temporal filtering (440a, 440b, 440c) to each of the decoded at least two sub-band signals;

spatially recomposing (450) the at least two sub-band signals; and

reconstructing the video signal from at least one of the at least two spatially recomposed sub-band signals.

24. The method according to claim 23, wherein the video signal is reconstructed from all of the at least two spatially recomposed sub-band signals.

25. A memory medium for decoding video, the memory medium comprising:

code for decoding a signal (420) including at least two encoded motion compensated temporally filtered, different frequency sub-band signals of a video signal;

code for independently applying inverse motion compensated temporal filtering (440a, 440b, 440c) to each of the decoded at least two sub-band signals;

code for spatially recomposing (450) the at least two sub-band signals; and

code for reconstructing the video signal from at least one of the at least two spatially recomposed sub-band signals.

26. A device for decoding video, the device comprising:

a texture decoding unit (420) for decoding a signal including at least two encoded motion compensated temporally filtered, different frequency sub-band signals of a video signal;

an inverse motion compensated temporal filtering unit (440a, 440b, 440c) for each of the at least two sub-band signals, each inverse motion compensated temporal filtering unit independently applying inverse motion compensated temporal filtering to its associated decoded at least two sub-band signal;

an inverse wavelet transform unit (450) for spatially recomposing the at least two sub-band signals; and

a video reconstructing unit for reconstructing the video signal from at least one of the at least two spatially recomposed sub-band signals.

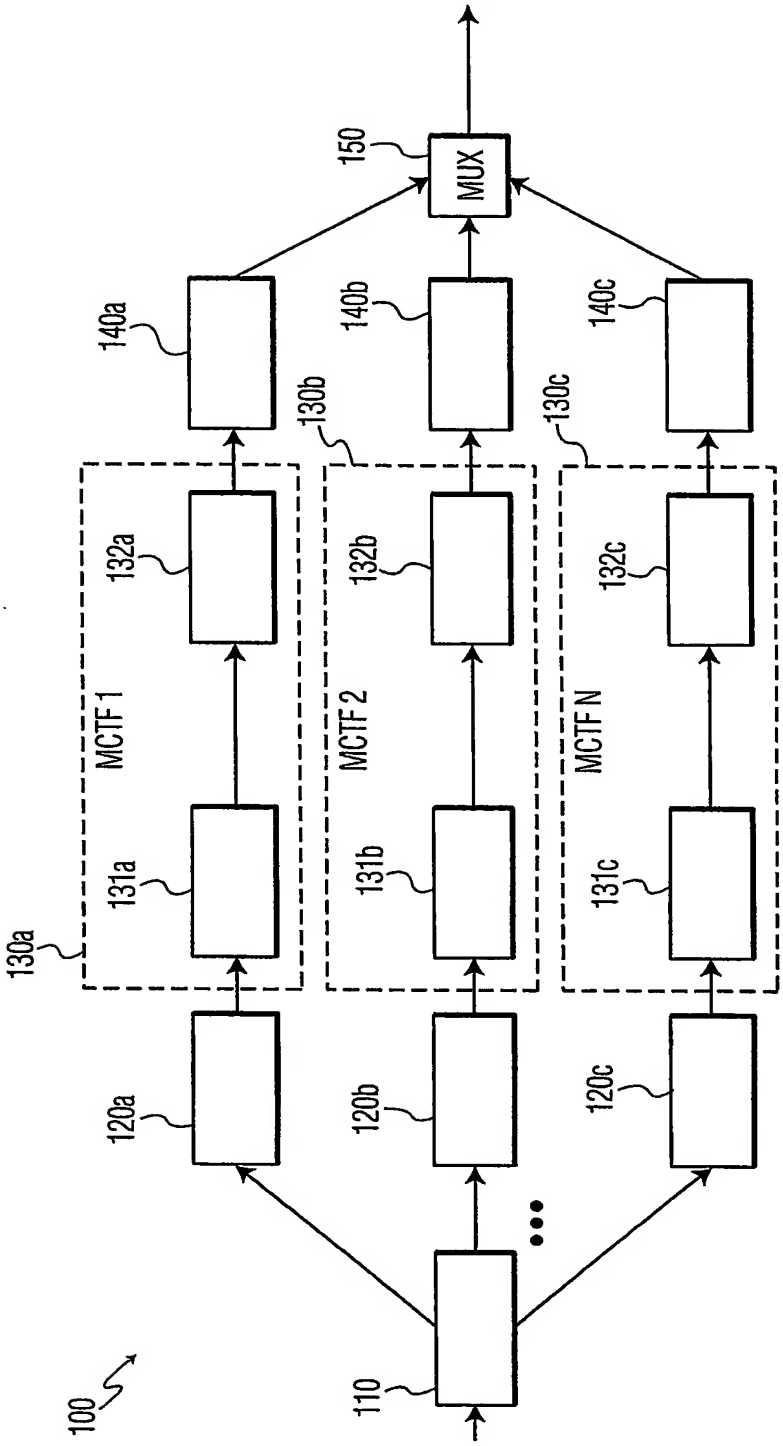

100
110
120a
120b
120c
130a
130b
130c
131a
131b
131c
132a
132b
132c
140a
140b
140c
150
MUX
MCTF 1
MCTF 2
MCTF N

FIG. 1

FIG. 2

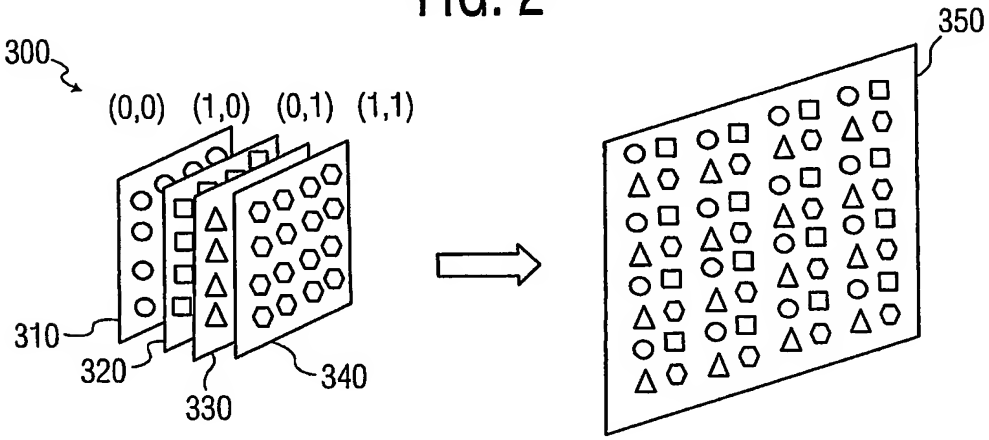


FIG. 3

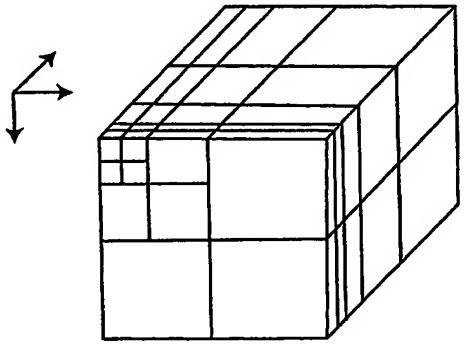


FIG. 4A

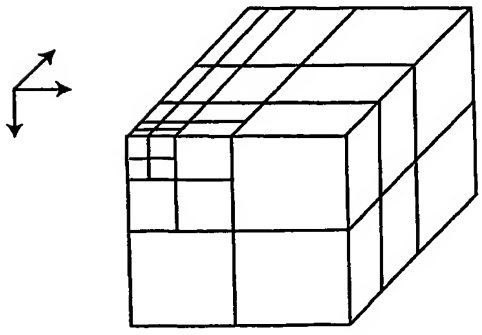


FIG. 4B

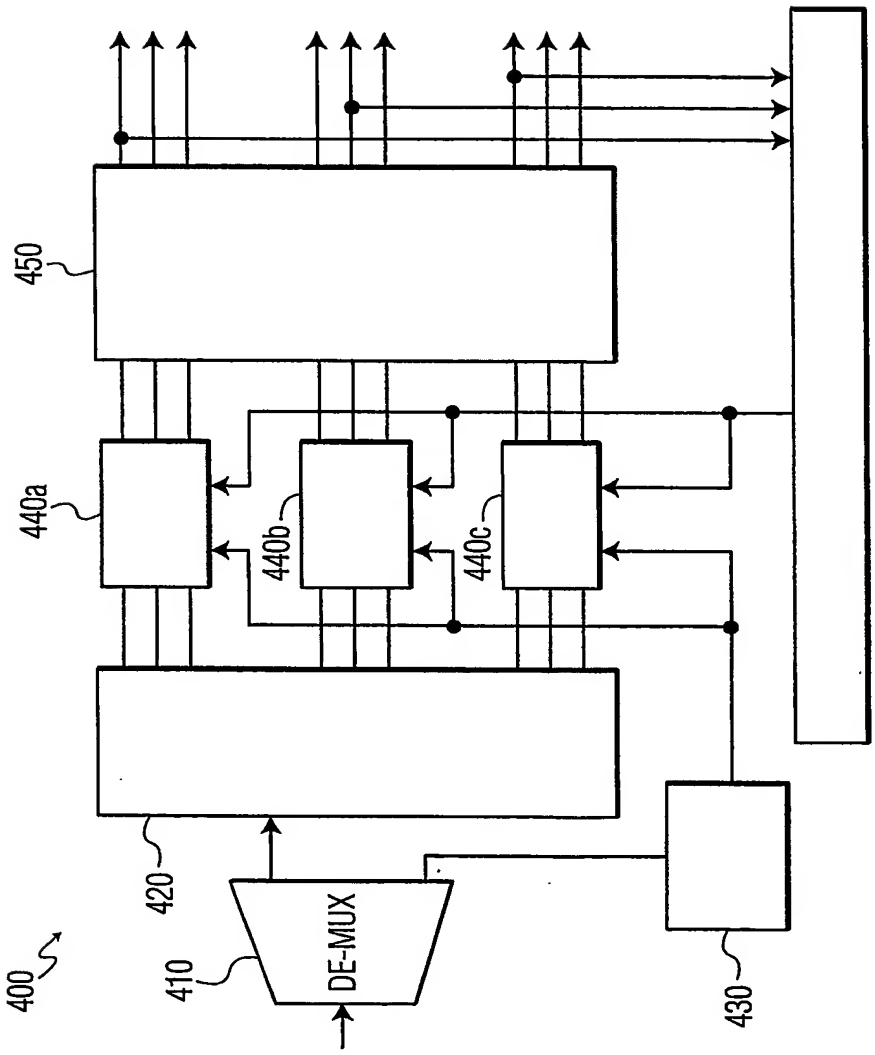


FIG. 5

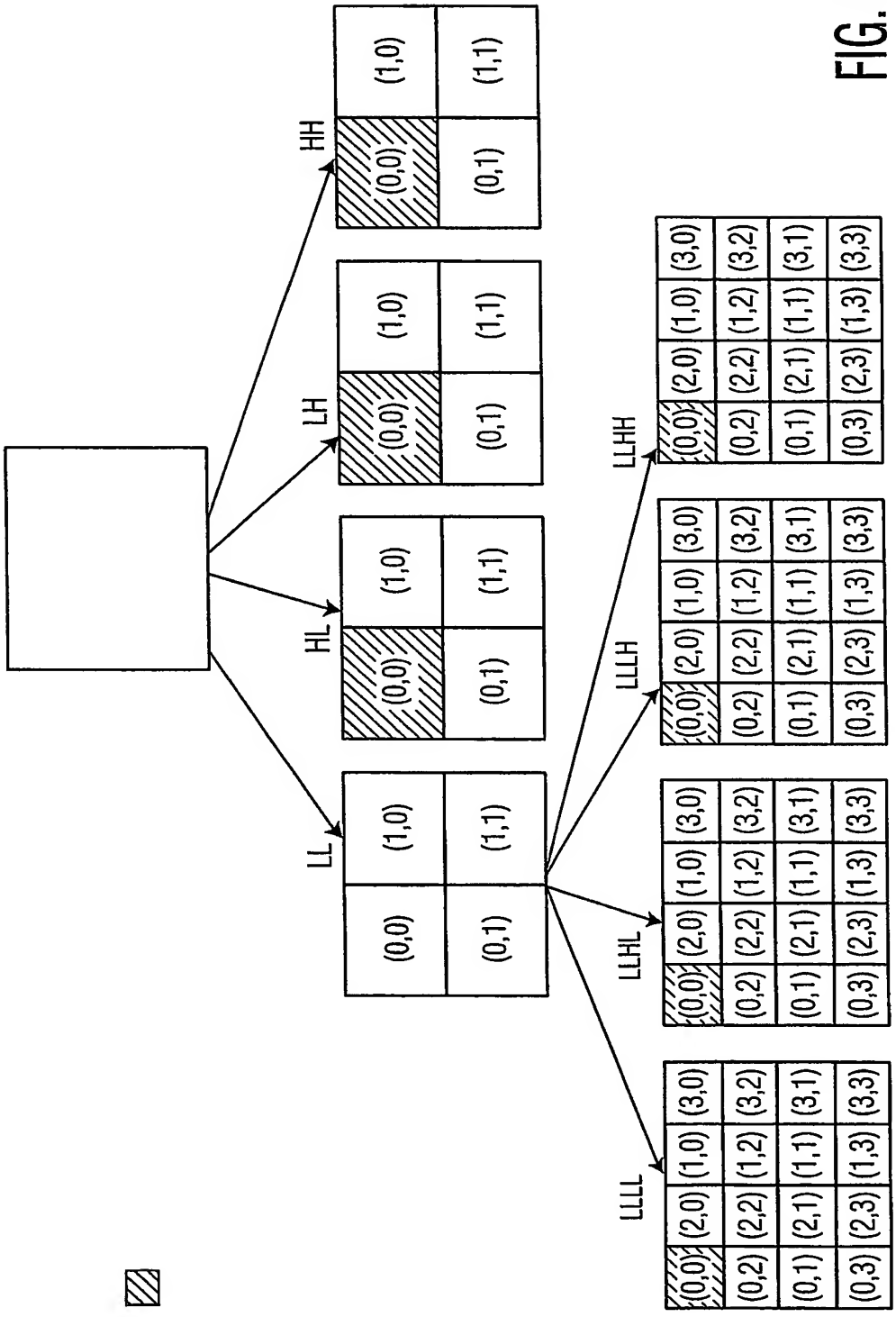

LL
HL
LH
HH
LLLL
LLHL
LLLH
LLHH
(0,0)
(0,1)
(1,0)
(1,1)
(0,2)
(0,3)
(1,2)
(1,3)
(2,0)
(2,1)
(2,2)
(2,3)
(3,0)
(3,1)
(3,2)
(3,3)

FIG. 6

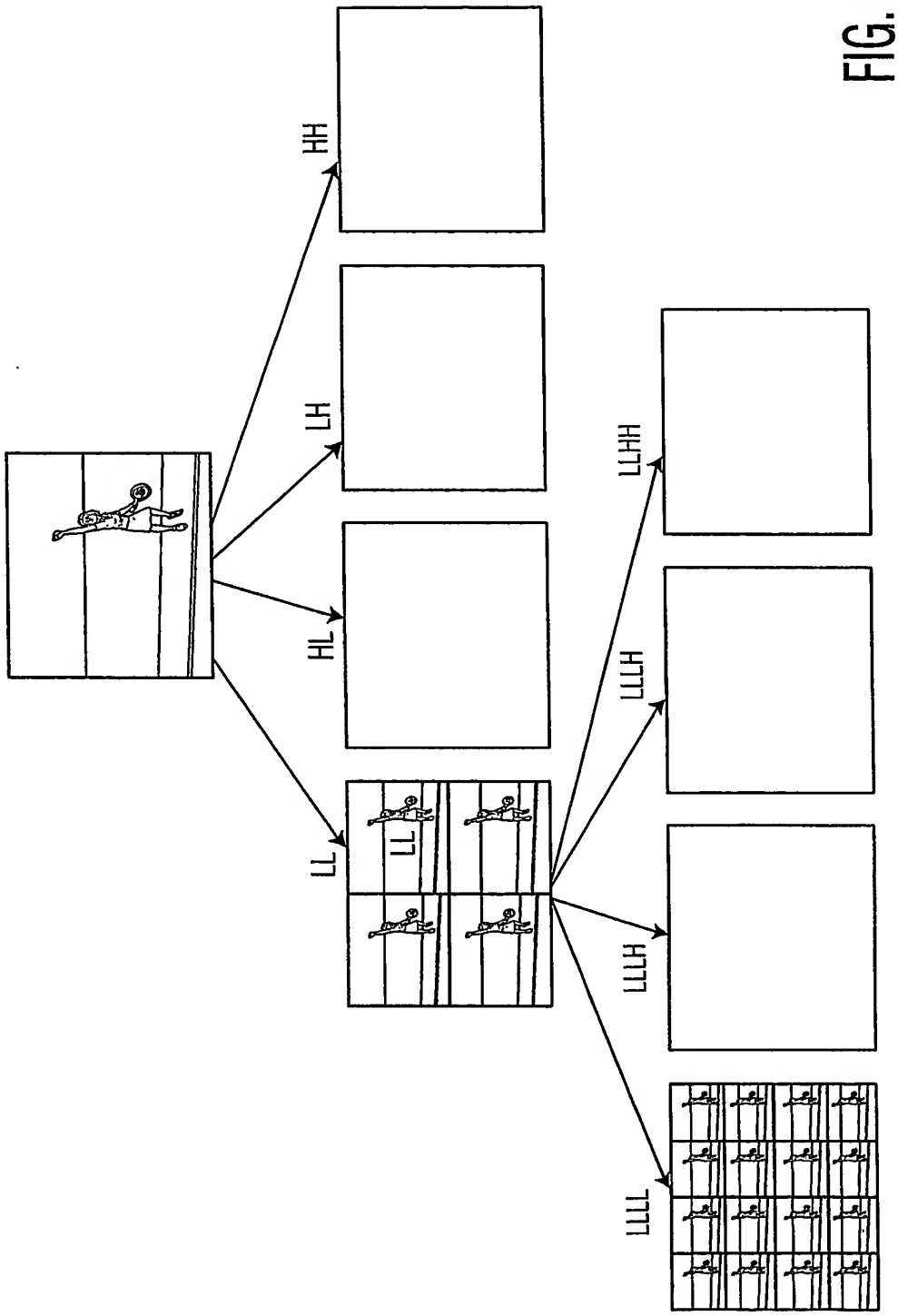


FIG. 7

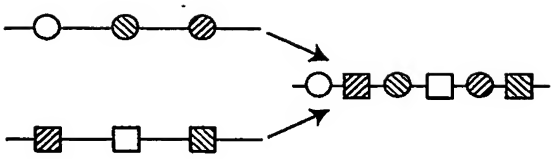


FIG. 8

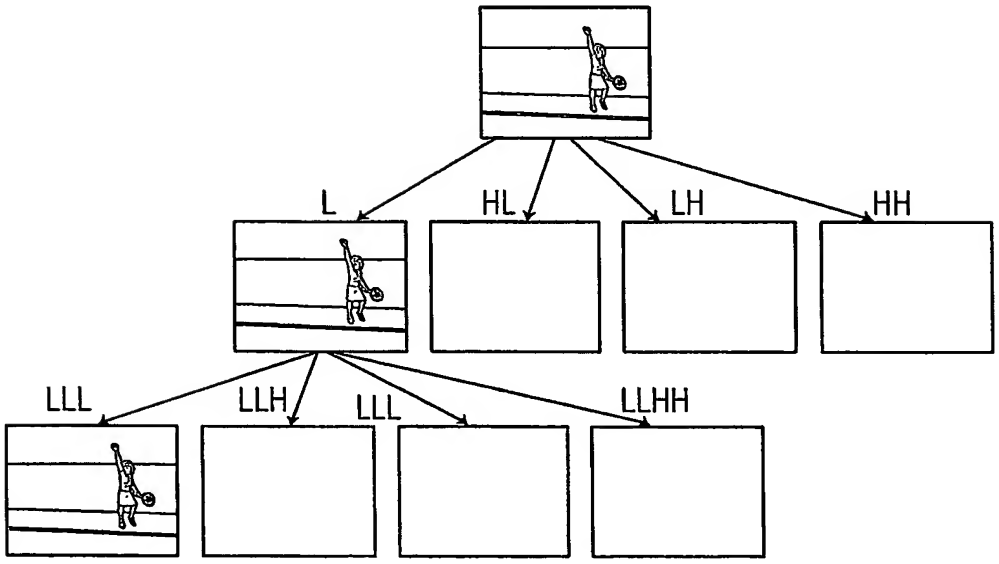


FIG. 9

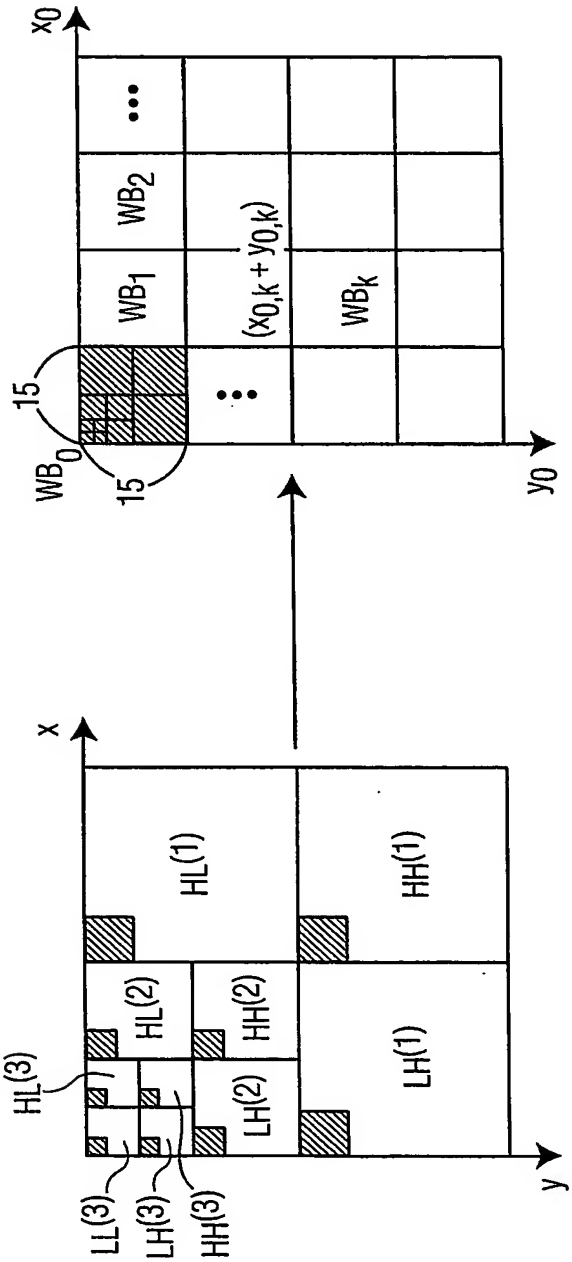

x0
y0
WB0
WB1
WB2
WBk
(x0,k + y0,k)
15
15
x
y
HL(1)
HH(1)
HL(2)
HH(2)
LH(1)
LH(2)
HL(3)
LL(3)
LH(3)
HH(3)

FIG. 10

		SDMCTF	IMBCTF
		4.0825	4.0143
		3.3402	3.2732
		1.8435	1.8414
		1.6400	1.6090
		5.7399	5.7214
		4.1163	4.1035

FIG. 11

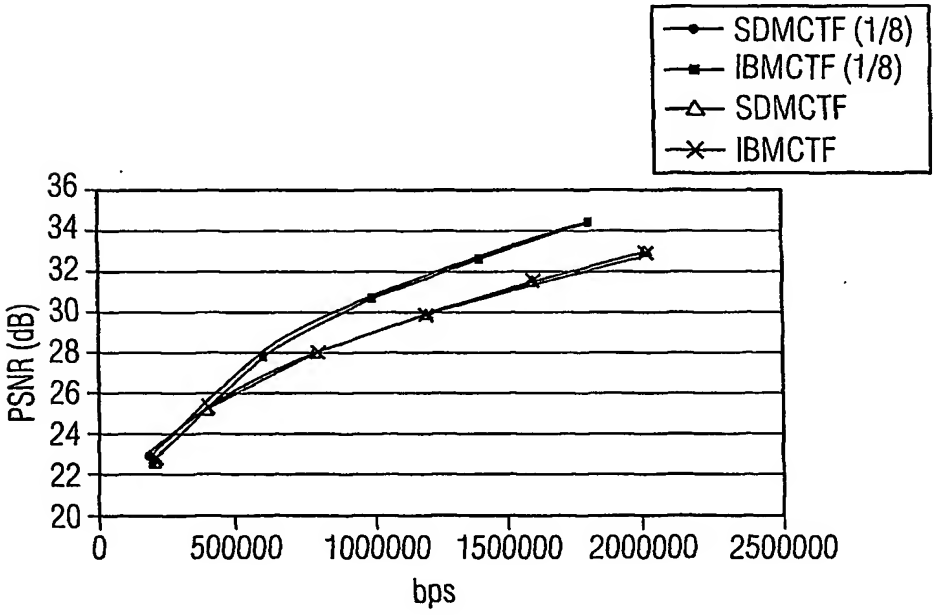


FIG. 12

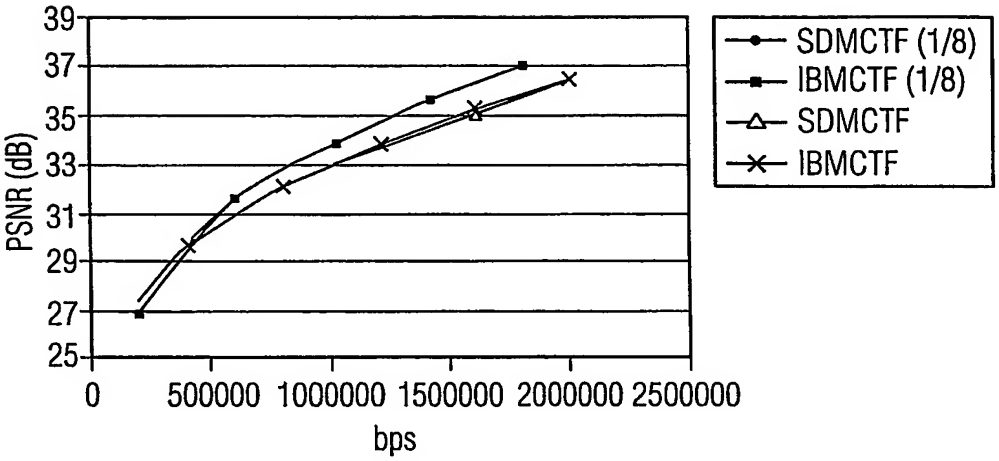

39
37
35
33
31
29
27
25
PSNR (dB)
0
500000
1000000
1500000
2000000
2500000
bps
SDMCTF (1/8)
IBMCTF (1/8)
SDMCTF
IBMCTF

FIG. 13

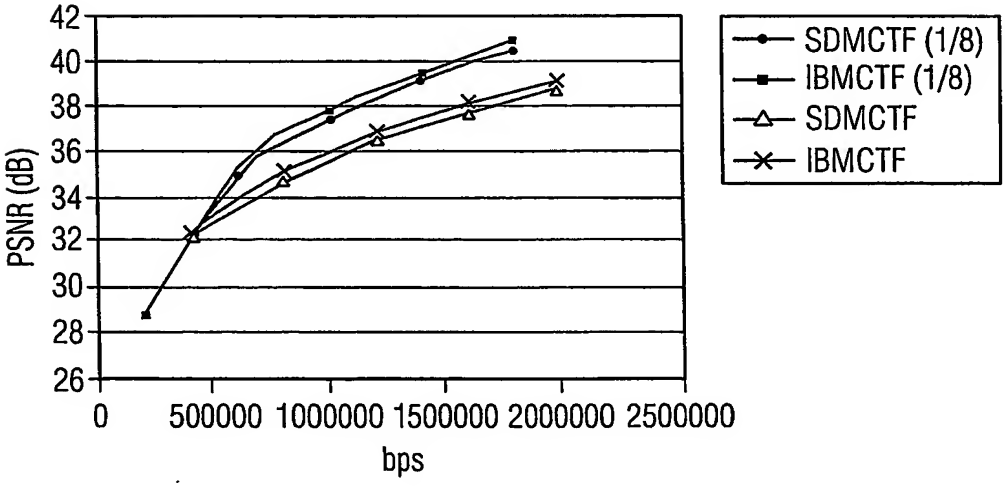

42
40
38
36
34
32
30
28
26
PSNR (dB)
0
500000
1000000
1500000
2000000
2500000
bps
SDMCTF (1/8)
IBMCTF (1/8)
SDMCTF
IBMCTF

FIG. 14

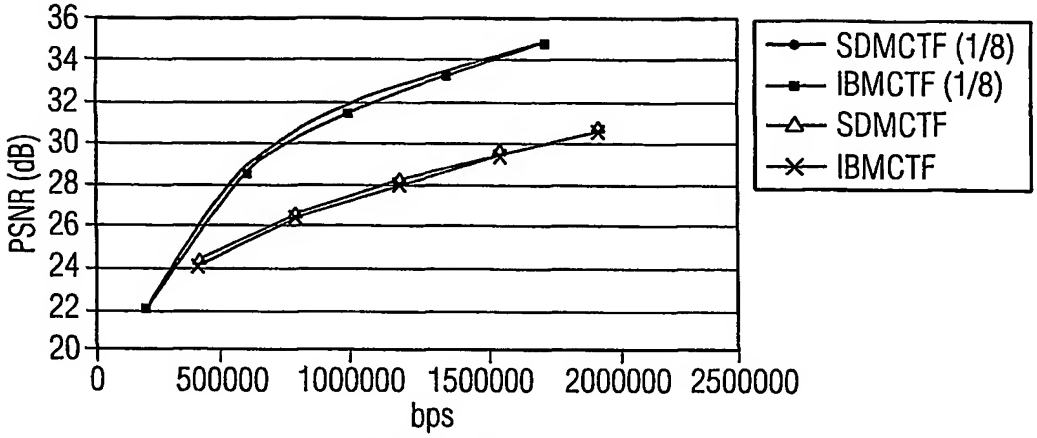


FIG. 15

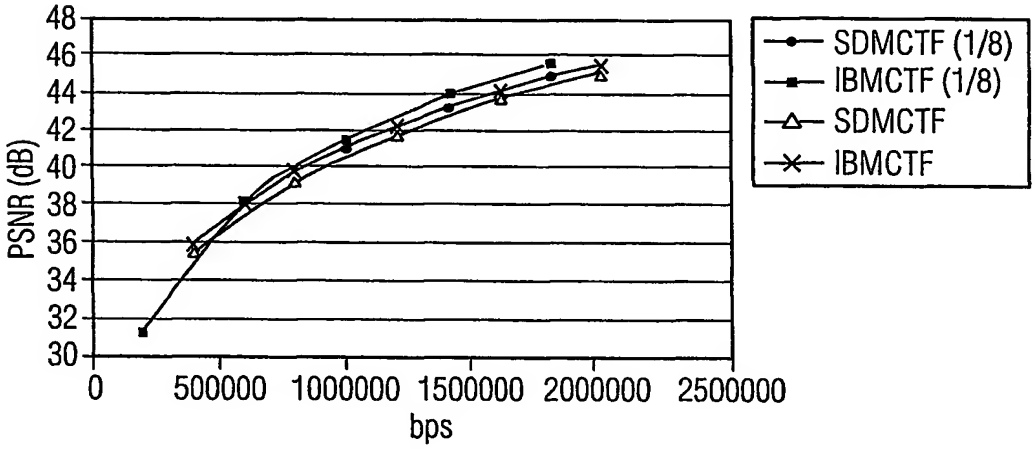


FIG. 16

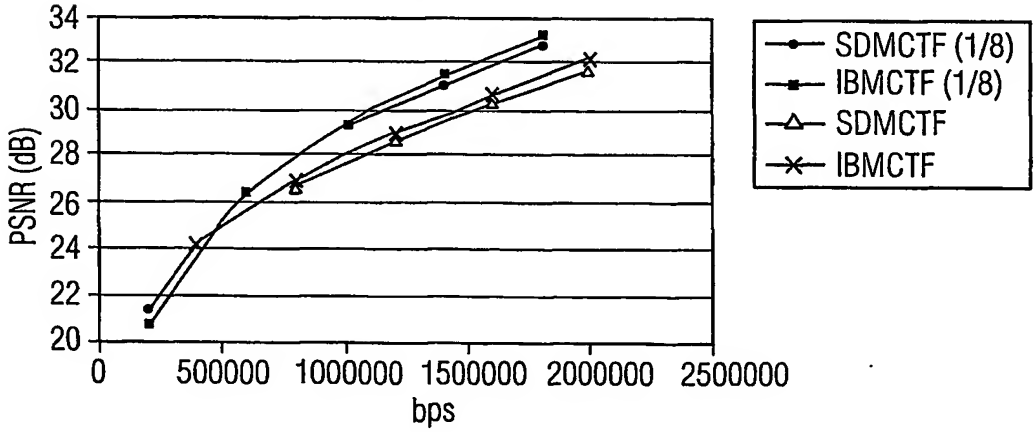


FIG. 17

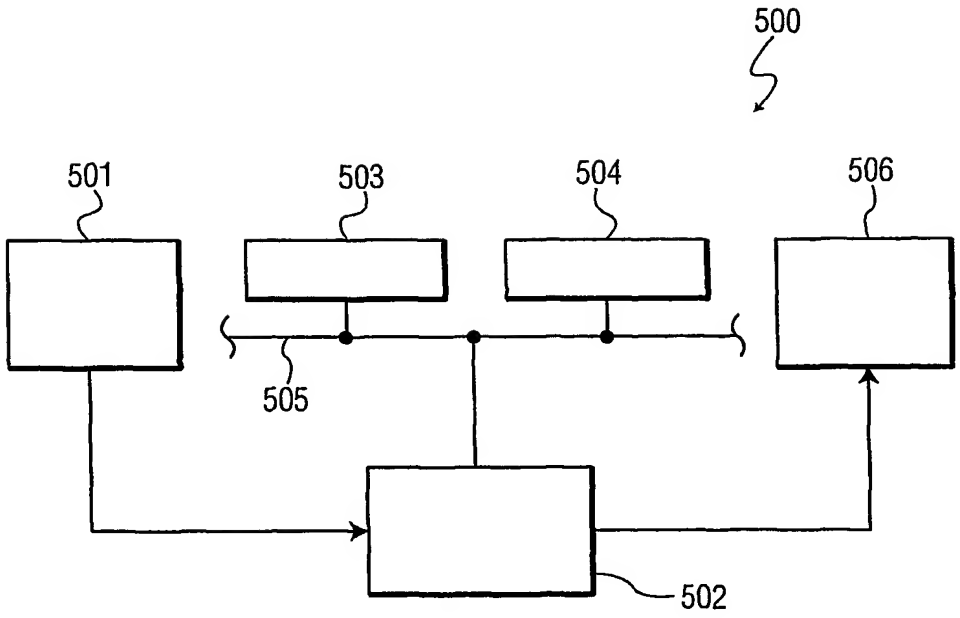


FIG. 18

INTERNATIONAL SEARCH REPORT

Application No
PCT/IB 03/04452

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 H04N7/26

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 H04N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, WPI Data, COMPENDEX, INSPEC, IBM-TDB

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	LI XIN ET AL: "Efficient motion field representation in the wavelet domain for video compression" INTERNATIONAL CONFERENCE ON IMAGE PROCESSING (ICIP'02);ROCHESTER, NY, UNITED STATES SEP 22-25 2002, vol. 3, 2002, pages III/257-III/260, XP002268097 IEEE Int Conf Image Process;IEEE International Conference on Image Processing 2002	1,2,7,8, 10,11, 15,17, 18,23-26
A	page 257, right-hand column, line 16 - line 24 page 258, left-hand column, paragraph 1 - paragraph 2; figure 1 page 259, left-hand column, last paragraph -right-hand column, last paragraph; figure 4	3-6,9, 12-14, 16,19-22

-/--

[X] Further documents are listed in the continuation of box C. [ ] Patent family members are listed in annex.

° Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
30 January 2004	24/02/2004

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31.651 epo nl, Fax: (+31-70) 340-3016	Heising, G

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
X	Y. ANDREOPOULOS, A. MUNTEANU, G. VAN DER AUWERA, P. SCHELKENS AND JAN CORNELIS: "Wavelet-Based Fully-Scalable Video Coding with In-Band Prediction" 3RD IEEE BENELUX SIGNAL PROCESSING SYMPOSIUM (SPS-2002), 21 - 22 March 2002, pages S02-1-S02-4, XP002268099 Leuven, Belgium	1,2,12, 17,18, 23-26
Y	page 1, right-hand column, line 1 -page 3, left-hand column, line 23	3-6,9, 13,14, 16,19-22
A	figures 1-3	7,8,10, 11,15
Y	PEISONG CHEN AND JOHN W. WOODS: "Comparison of MC-EZBC and H.26L TML 8 on Digital Cinema Test Sequences" ISO/IEC JTC1/SC29/WG11 MPEG2002/M8130, 11 - 15 March 2002, pages 1-6, XP002268096 Jeju Island, Korea	3-6,9,19
A	page 1, paragraph 1 -page 2, last paragraph; figure 1	1,2,7,8, 10-18, 20-26
Y	VAN DER AUWERA G ET AL: "Scalable wavelet video-coding with in-band prediction - The bottom-up overcomplete discrete wavelet transform" PROCEEDINGS 2002 INTERNATIONAL CONFERENCE ON IMAGE PROCESSING. ICIP 2002. ROCHESTER, NY, SEPT. 22 - 25, 2002, INTERNATIONAL CONFERENCE ON IMAGE PROCESSING, NEW YORK, NY: IEEE, US, vol. 3 OF 3, 22 September 2002 (2002-09-22), pages 725-728, XP010607820 ISBN: 0-7803-7622-6	13,14, 16,20-22
A	page 725, left-hand column, paragraph 2 -page 727, right-hand column, paragraph 2; figure 1	1-12,15, 17-19, 23-26
A	THOMAS RUSERT AND MATHIAS WIEN: "Exploration Experiments on Spatial and Temporal Scalability in Interframe Wavelet Coding" ISO/IEC JTC1/SC29/WG11 MPEG2002/M8650, 22 - 26 July 2002, pages 1-7, XP002268098 Klagenfurt, Austria page 5, paragraph 1 -page 6, last paragraph; figure 2	1-26

-/--

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	D.S. TURAGA AND M. VAN DER SCHAAR: "Unconstrained motion compensated temporal filtering" ISO/IEC JTC1/SC29/WG11 MPEG2002/M8388, 6 - 10 May 2002, pages 1-15, XP002268488 Fairfax, US page 4, line 16 -page 7, line 21 ---	1-26
A	HSIANG S T ET AL: "Embedded video coding using invertible motion compensated 3-D subband/wavelet filter bank" SIGNAL PROCESSING. IMAGE COMMUNICATION, ELSEVIER SCIENCE PUBLISHERS, AMSTERDAM, NL, vol. 16, no. 8, May 2001 (2001-05), pages 705-724, XP004249801 ISSN: 0923-5965 page 709, paragraph 2 -page 713, paragraph 1; figures 3-6 -----	1-26